应用光学
Journal of Applied Optics
SCOPUS,等来源期刊
中文核心期刊、中国科技核心期刊

基于卷积神经网络的路面裂缝分割设计与研究

刘艳宁 章国宝

Design and research on pavement crack segmentation based on convolutional neural network

LIU Yanning, ZHANG Guobao

引用本文:
刘艳宁, 章国宝. 基于卷积神经网络的路面裂缝分割设计与研究[J]. 应用光学, 2024, 45(2): 373-384. DOI: 10.5768/JAO202445.0202004
LIU Yanning, ZHANG Guobao. Design and research on pavement crack segmentation based on convolutional neural network[J]. Journal of Applied Optics, 2024, 45(2): 373-384. DOI: 10.5768/JAO202445.0202004

在线阅读 View online: https://doi.org/10.5768/JAO202445.0202004

文章编号：1002-2082 (2024) 02-0373-12

基于卷积神经网络的路面裂缝分割设计与研究

刘艳宁，章国宝
（东南大学 自动化学院，南京 210000）

摘　要：裂缝是路面病害最主要的类型，准确的裂缝分割是国家进行公路预防养护管理的重要决策依据。针对背景复杂下现有模型路面裂缝分割准确度有待提高的问题，提出一种基于卷积神经网络的端到端裂缝分割模型，使用分层结构的 ConvNeXt 编码器提取多尺度特征，特征的最高层使用金字塔池化模块进一步获取全局先验特征，通过具有横向连接和自上而下的金字塔结构进行特征融合。针对裂缝和背景不平衡问题，使用平衡交叉熵损失函数提高模型的检测性能。此外，构建了一个包含 2876 张裂缝图片的数据集 UCrack，覆盖多种裂缝类型和广泛的背景范围，以提供丰富的特征供模型学习。实验表明，在 UCrack 测试数据集上模型的召回率和 *F*1 得分比其他表现最佳的模型提高了 2.68% 和 6.89%；在 CrackDataset 数据集上的测试取得了 85.68% 的召回率和 80.11% 的 *F*1 得分，说明模型具有较好的泛化性能，可应对背景复杂的路面裂缝分割。

关键词：裂缝分割；卷积神经网络；编解码网络；特征金字塔；金字塔池化
中图分类号：TN206;TP391　　**文献标志码：**A　　**DOI:** 10.5768/JAO202445.0202004

Design and research on pavement crack segmentation based on convolutional neural network

LIU Yanning, ZHANG Guobao
（School of Automation, Southeast University, Nanjing 210000, China）

Abstract: Cracks are the most important type of pavement diseases, and the accurate crack segmentation is an important decision basis for national preventive maintenance management of roads. To address the problem of crack segmentation accuracy of existing models for pavement under complex background, an end-to-end crack segmentation model based on convolutional neural network was proposed, which used a layered structure of ConvNeXt encoder to extract multi-scale features. A pyramid pooling module was used to further obtain the global priori features by the top layer of features, and the feature fusion was performed through a pyramid structure with lateral connections and top-down. A weighted cross-entropy loss function was employed to enhance the detection performance of model for the crack and background imbalance problem. In addition, a crack dataset UCrack with 2 876 cracks covering multiple crack types and a wide range of backgrounds was created to provide rich features for model learning. Experiments show that, compared with other best-performing models, the model recall and *F*1 score on the UCrack test dataset are improved by 2.68% and 6.89%, respectively. The test on the CrackDataset dataset achieves recall of 85.68% and *F*1 score of 80.11%, which implies that the model has better generalization capability and can cope with pavement crack segmentation with complicated scenarios.

Key words: crack segmentation; convolutional neural network; encoder-decoder network; feature pyramid; pyramid pooling

收稿日期：2023-04-23；　**修回日期：**2023-08-17
基金项目：江苏省重点研发计划 (BE2020116, BE2021750)
作者简介：刘艳宁（2000—），女，硕士研究生，主要从事计算机视觉研究。E-mail：lyn_129@163.com
通信作者：章国宝（1965—），男，教授，博导，主要从事机器学习、深度学习和计算视觉研究。E-mail：zhangguobao65@163.com

引言

预防性道路养护是提高路面使用性能和延长路面使用寿命的最有效手段，裂缝作为路面病害最常见、最易发生和最早期产生的类型，若任其发展，雨雪水会通过裂缝渗入基层和土基，降低路基的稳定性和强度，引发其他类型严重的路面病害，对行驶车辆和整个道路交通造成严重影响。因此，进行路面裂缝分割以获取准确的路面裂缝信息，为道路预防性养护提供决策依据，是目前道路养护工作的前提[1]。路面裂缝分割任务主要面临着以下挑战：其一，路面裂缝的形态结构复杂，形状、大小和方向均不固定，尺度多变；其次，裂缝图像背景复杂，除了交通标志线和路面纹理差别产生的干扰，还容易受自然环境的影响，如异物遮挡、光线差异和油污水渍都会干扰裂缝分割任务的准确性。

近年来，随着深度学习卷积神经网络 AlexNet[2] 在 2012 年 ImageNet 图像分类竞赛[3] 中大获全胜，基于深度学习的研究和应用在各个领域发展，国内外基于神经网络的路面裂缝分割研究也不断增多。YANG X 等[4] 使用全卷积网络解决裂缝分割问题，实现了端到端的像素级裂缝分割，验证了全卷积网络可应用于裂缝识别，但是测量精度较低。语义损伤检测网络[5] 提出基于改进的空洞空间卷积池化金字塔的卷积神经网络，提高了计算效率，但是该方法的数据集只有 200 张单一类型的裂缝图像，没有验证模型的泛化性能。于海洋等[6] 基于残差和注意力机制改进 U-Net 网络，改善了背景噪声干扰导致的检测精度低问题，但是使用的数据集有限。JI A 等[7] 将语义分割网络 DeepLabv3+[8] 的集成方法用于裂缝分割，并提出了一种用于像素级裂缝量化的量化算法，但是受小裂缝和噪声干扰的影响，在各种实际场景中的泛化能力有待提升。廖延娜[9] 等提出基于 Mask RCNN 的裂缝检测算法，针对裂缝形态复杂多样，将裂缝划分为裂缝和破损两类别，针对性地提高了网络在破损类裂缝检测的准确度，但需要额外的分类标注，提高了人工成本。

上述基于深度学习的方法取得良好的效果，但在实际应用场景下的分割有待进一步研究。一方面，裂缝自身的形态结构复杂多样，路面裂缝分割模型对于统一多尺度裂缝检测准确度有待提高；另一方面，所使用的数据集大多是干净无污渍、清晰无遮挡的图像，数据集覆盖范围和大小有限，缺乏基于复杂背景裂缝数据集的研究。针对上述原因，本文基于编解码的金字塔网络 (feature pyramid networks, FPN)[10] 结构设计裂缝分割模型，特征金字塔网络结构通过横向连接和自上而下的结构，将多层次的特征图进行多分支融合，以保留各层次特征图的语义信息。为有效提取背景复杂下的多尺度裂缝特征，提高实际应用场景的检测准确度，本文主要有以下改进：编码器引入可以从图像中提取更详细特征的 ConvNeXt[11] 模块，结合金字塔的分层结构能够有效提取多尺度特征，缓解编码器在特征提取过程中由于感受野扩大而导致细小裂缝特征信息丢失的问题；特征图的最高层接入金字塔池化模块，减少由于高层特征图的多通道融合导致的上下文丢失信息，充分利用全局信息以针对性地改善多尺度裂缝的漏检问题；最后为减小解码过程中由于上采样导致细节丢失问题，使用密集上采样模块替换双线性插值模块。经实验验证，本文提出的基于卷积神经网络的路面裂缝分割模型，与其他 4 种语义分割方案相比，表现出更佳的路面裂缝分割鲁棒性。此外，针对上述原因二，为满足验证模型在复杂背景下分割的性能对数据集覆盖度和大小的需求，本文构建了一个包含 2876 张裂缝图像的大型数据集，广泛覆盖了不同的路面类型、裂缝类型、光照强度和噪声干扰。

1 基于分层结构的路面裂缝分割模型

本文设计的基于分层结构的路面裂缝分割模型如图 1 所示。整体设计基于特征金字塔网络，分为 3 个主要结构：一是自下而上的编码部分，编码器进行多尺度特征提取，生成从高分辨率粗特征到低分辨率细特征的 4 层特征图；二是自上而下的解码部分，对语义信息更丰富的高层特征进行逐级上采样，使得低层特征也包含丰富的语义信息；并通过第 3 种横向结构，实现编码和解码部分相同层的特征融合，得到同时包含低级和高级语义信息的特征映射集，分别相当于原始输入图像 1/4、1/8、1/16、1/32 的分辨率。最后，通过卷积层消除上采样产生的混叠效应，通过双线性插值将不同层的特征图变换到相同尺度，并由卷积层完成不同层特征图的融合，以进行路面裂缝的像素级分类。

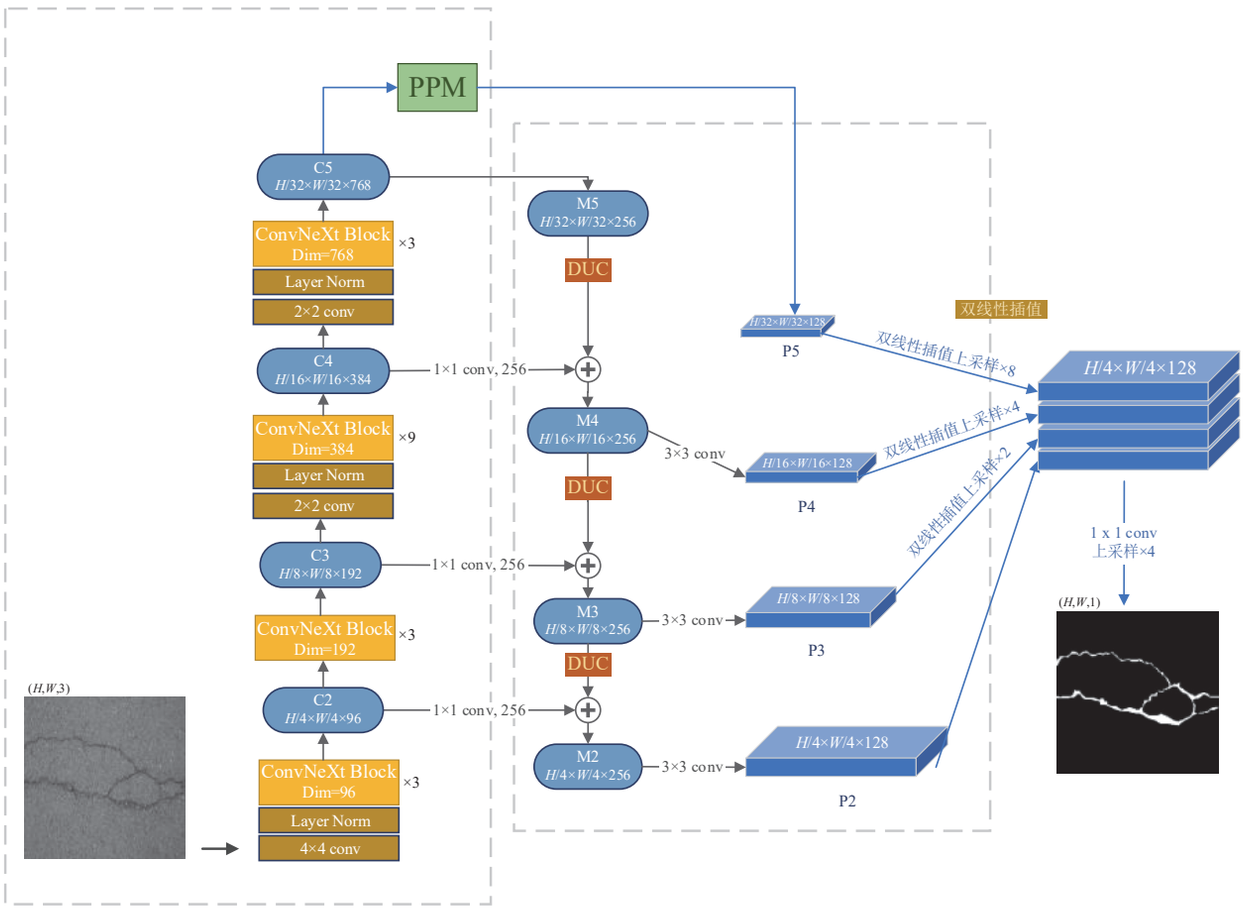

图 1　路面裂缝分割算法流程

Fig. 1　Algorithm process of pavement crack segmentation

1.1　ConvNeXt 模块

近年 Swin-Transformer[12] 在图像分类、图像分割领域表现出强大的性能优势; 而 ConvNeXt 基于残差神经网络[13], 引入 MobileNet[14]、Vit[15]、Swin-Transformer[12] 等网络的思想和特点进行改进, 在图像任务中表现出更快速度和更高精度的优势。如图 2 所示, ConvNeXt 模块由深度可分离卷积[16]、层归一化[17]、传统卷积和 GELU 激活函数层组成。7×7 大小的深度可分离卷积用于提取第 1 层的空间信息, 与传统卷积相比, 深度可分离卷积的一个卷积核只与一个通道进行计算, 具有参数数量较少和计算量较小的优点, 与 Swin Transformer 相比, ConvNeXt 将深度可分离卷积上移至第 1 层, 并验证了在具有相同的卷积核大小前提下, 其有助于显著减少浮点运算。在激活函数的使用设计上, ConvNeXt 模块借鉴 Transformer 的思想使用 GELU 激活函数, 发现 ConvNeXt 仅在两个 1×1 卷积层之间使用一层 GELU 激活函数, 而不是在每个卷积层都附加一个激活函数时, 能够一定程度上提高模型的精度。借鉴 Transformer 具有更少归一化层的设计, ConvNeXt 模块仅在第一个 1×1 卷积层之前添加归一化操作, 在不会降低精度的情况下, 降低了模型的参数量。整体设计上, ConvNeXt 中间

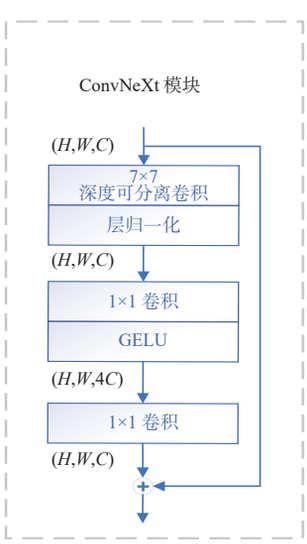

图 2　ConvNeXt 模块[11]

Fig. 2　ConvNeXt module

层通道数为输入和输出层通道数的 4 倍，这种小维度-大维度-小维度的反瓶颈结构，能够避免不同维度特征空间之间转换时由于压缩维度带来的信息损失。考虑 ConvNeXt 在图像任务中的优异表现，本文模型在编码器中引入 ConvNeXt 模块替换普通残差模块来提取特征，结合分层结构生成特征图集，根据多尺度特征图的融合结果进行像素级别的裂缝元素分类，并验证 ConvNeXt 在裂缝图像的像素级分类领域的优秀分类效果。

1.2 密集上采样 DUC 模块

图像分割任务使用上采样解码来提高特征图的分辨率，将特征图恢复到原图大小。分割模型如 DeepLabv3+，通常采用双线性插值进行上采样，但是此方法只考虑待测像素点的直接邻点灰度值，而不考虑灰度值变化率的影响，导致图像边缘在一定程度上变模糊，丢失大量细节信息，这对裂缝分割任务的边缘分割非常不利。密集上采样 DUC（dense upsampling convolution）模块[18] 直接在特征图上应用卷积操作，与双线性上采样相比，能够解码更细节的信息，被证明尤其在相对较小的物体上效果更好，因此更适合裂缝这种细长类别的目标。对于输入大小为$(H/r,W/r,C)$的特征图，想要恢复到(H,W,C)的大小，如图 3 所示，首先进行 1×1 卷积生成$(H/r,W/r,r^2\times C)$维的输出特征图，这样划分为r^2个$(H/r,W/r,l)$大小相同的子特征图，然后使用 softmax 层将其叠加映射为 $(H, W, 1)$ 大小的特征图。这种上采样方式，使得每一层都可以学习对像素的预测，提高了模型的学习能力，降低了细节损失。

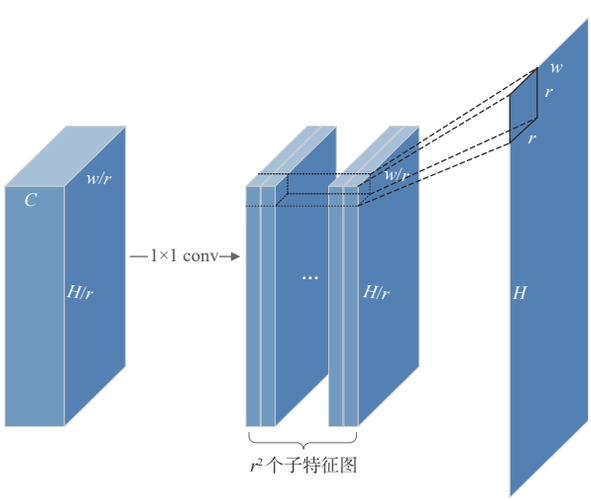

图 3 DUC 模块[18]

Fig. 3 DUC module

1.3 金字塔池化模块

充分利用上下文信息对于分割任务有明显影响，UperNet[19] 发现 FPN 网络与能够带来有效全局先验信息的金字塔池化模块 (pyramid pooling module, PPM)[20] 高度兼容，可以减少不同区域间上下文信息的丢失，提高获取全局信息的能力。如图 4 所示，PPM 通过不同尺度的池化操作和一系列卷积操作获取 4 个不同大小的特征图来扩大感受野，然后对获取到的特征图进行上采样至和原始输入特征图相同的大小，将其与原输入特征图拼接在一起组成最终的特征表达。通过 PPM 模块，能进一步从特征图的最高层获取全局的场景信息，并将其送入解码模块的顶层作为先验，这种基于全局的场景解析，获取了更充分的场景上下文信息，非常有利于裂缝这种多尺度的分割任务。

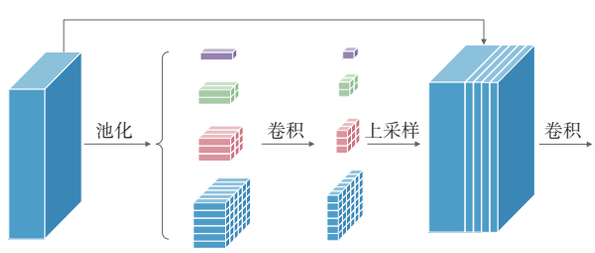

图 4 金字塔池化模块[20]

Fig. 4 Pyramid pooling module

2 实验过程

2.1 构建数据集

近年来许多学者开源了自己开发的裂缝分割数据集，并手工进行了像素级标注。考虑现实应用场景的复杂性，通用的语义级分割裂缝数据集通常需要涵盖非常广泛的路面场景、裂缝尺度和噪声类型。对于完好的路面场景，背景干净纹理光滑，裂缝和背景对比明显；对于受损的路面场景，常常伴有坑洼和凸起的粗糙表面；对于脏乱的路面场景，则分布大量的油污斑点等污渍，可能还存在其他道路垃圾的干扰。同时，拍摄路面裂缝数据时，往往存在光线不足、光照不均和自然光投射产生的阴影等问题。除此之外，裂缝的形态结构也复杂多变。然而目前开源的数据集往往局限于单一类型，如 CFD 数据集只包含完好的沥青路面一种场景，覆盖的噪声类型和裂缝尺度非常有限。若数据集的变化较小，训练的裂缝分割模型容易存在过拟合问题。针对以上问题，本文将现有的 8 个数据集合并为 1 个涵盖范围广的完整数

据集, 称为 UCrack, 用于本文裂缝分割研究。如表 1 所示, 共有 2876 张图片, 涵盖了各种路面场景, 具有较为广泛的照片亮度、噪声干扰和裂缝尺度分布, 可以较好地解决模型过拟合问题。

表 1　所构建的 UCrack 数据集概况

Table 1　Overview of established UCrack dataset

数据集名称	数量	尺度/像素	数据集特点描述
ESAR[21]	15	768×512	自然状态下动态拍摄的沥青路面; 包含阴影干扰问题
AIGLE_RN[21]	38	991×462 311×462	动态拍摄的沥青路面; 存在严重的光照不均和背景椒盐噪点问题
DeepCrack[22]	300	544 ×384	混泥土和沥青路面; 包含完好路面、受损路面、脏乱路面等场景
GAPs384[23]	509	640×540	德国沥青路面; 包含坑洼, 斑块等干扰问题
Crack500[24]	1 896	640×340	美国天普大学主校区的混泥土和沥青路面; 包含车道线、坑洼破损、椒盐噪声等强干扰问题
CFD[25]	118	480×320	中国北京的沥青路面状况; 包含阴影、车道线、油斑和水渍等干扰问题

此外, 为了验证模型的泛化性能, 本文的泛化性测试数据集有 3 部分, 一部分是作者在南京市将军大道附近拍摄的 156 张路面图像, 另外是公开数据集 CrackDataset[4] 和 CRKWH100[26], 二者均未参与模型训练, 仅用于泛化性测试。其中 CrackDataset 共有 776 张裂缝图片, 随机划分出 4/5 的数据作为模型的泛化性能测试数据集, 剩下 1/5 作为 3.2 节对比实验的训练数据, 不影响泛化性测试。CRKWH100 共有 100 张裂缝图片, 注释信息仅为裂缝骨架, 所以仅用于定性实验。

为了提高训练速度, 方便后续处理, 本文将训练数据缩放至 448×448 像素大小, 注释文件转换为单通道模式保存, 裂缝像素点标签值为 1, 背景像素点标签值为 0。随机划分 60% 的数据作为模型的训练数据, 20% 作为验证数据, 用于优化模型, 并在训练过程中选择最佳模型, 20% 作为测试数据评估模型。为进一步增强模型的泛化能力, 在原始数据集上采用随机翻转、光度失真、随机噪声等手段进行数据增强。

2.2　评价指标

评价过程中, 所有模型均采用原比例单尺度预测方法, 采用以下评价指标来评估模型性能。

整体上, 平均像素准确度 (average Accuracy, aAcc, A_{aAcc}) 从像素级衡量整体的分类准确度, 平均准确度均值 (mean Accuracy, mAcc, A_{mAcc}) 从类别级衡量整体的分类准确度。针对目标裂缝, 裂缝精确率 (Precision, P) 是预测正确的裂缝像素在所有预测为裂缝像素中的占比, 召回率 (Recall, R) 是预测正确的裂缝像素在所有裂缝像素中的占比。以上评价指标公式如下:

$$A_{\mathrm{aAcc}}=\frac{T_{\mathrm{P}}+T_{\mathrm{N}}}{T_{\mathrm{P}}+T_{\mathrm{N}}+F_{\mathrm{P}}+F_{\mathrm{N}}} \tag{1}$$

$$A_{\mathrm{mAcc}}=\frac{1}{2}\left(\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}}+\frac{T_{\mathrm{N}}}{T_{\mathrm{N}}+F_{\mathrm{P}}}\right) \tag{2}$$

$$P=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}} \tag{3}$$

$$R=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}} \tag{4}$$

式中: T_{P}表示真阳性(实际裂缝像素被正确预测); T_{N}表示真阴性(实际非裂缝像素被正确预测); F_{P} 表示假阳性(实际非裂缝像素被错误预测为裂缝); F_{N}表示假阴性(实际裂缝像素被错误预测为非裂缝)。

精确率代表模型是否能准确预测裂缝元素, 精确度越高说明区分裂缝和噪声干扰的能力越强。召回率代表模型是否能检测出所有裂缝元素, 召回率越高, 说明裂缝元素越容易被正确检测。但是召回率越高, 一般代表模型越容易将元素预测为裂缝元素, 导致区分噪声干扰的能力变差, 所以精确率和召回率是一对需要衡量考虑的关联性指标。$F1$ 分数是精确率和召回率的调和指标:

$$F1=\frac{2\times P\times R}{P+R}=\frac{2T_{\mathrm{P}}}{2T_{\mathrm{P}}+F_{\mathrm{P}}+F_{\mathrm{N}}} \tag{5}$$

2.3　实验实施

2.3.1　实施细节

本文算法基于 Pytorch 框架 1.10 版本编写, 硬件平台为 32G 内存的 Intel(R) Core(TM) i9-9900KF, 并使用 NVIDIA GeForce RTX 2080 Ti 进行加速计算。

特征提取网络 ConvNeXt 使用预训练的模型进行初始化。采用 AdamW[27] 优化器, 权值衰减率设为 0.05, 训练批次设为 8, 在本文数据集上迭代 30000 次训练模型, 根据 $F1$ 分数挑选最佳模型作为训练结果。网络骨干部分的学习率初始值设置为 0.0001, 非骨干部分的学习率设置为其 10 倍,

以获取更好的性能和更快的收敛速度[28]。采用 poly 学习率衰减策略，在每个迭代周期动态调整学习率。

2.3.2 损失函数

裂缝分割是典型的样本不平衡分割任务，裂缝的像素远少于背景像素。使用交叉熵损失函数 (cross entropy, CE) 模型倾向于关注像素点更多的背景元素是否正确被预测，导致对裂缝元素的预测能力较差，容易发生误检现象，这与裂缝分割任务需要从图像复杂背景中准确地提取裂缝区域的目标相悖，其公式为

$$L_{\mathrm{CE}}=-\frac{1}{N}\sum_{0\leqslant i\leqslant N}[y_i\ln p_i+(1-y_i)\ln(1-p_i)] \tag{6}$$

式中：N为像素点总数；y_i和p_i分别代表第i个像素点的真实标签值和预测为裂缝像素点的概率值。

平衡交叉熵 (balanced cross entropy, BCE) 损失函数为损失函数添加权重因子，强制将注意力聚焦于较少样本[29-30]，其公式为

$$L_{\mathrm{BCE}}=-\frac{1}{N}\sum_{0\leqslant i\leqslant N}[w_1y_i\ln p_i+w_2(1-y_i)\ln(1-p_i)] \tag{7}$$

式中：w_1为裂缝像素点权重；w_2为背景像素点权重。

Focal 损失函数[31]是针对训练样本不平衡以及样本难易程度不同提出的损失函数，定义如下：

$$L_{\mathrm{Focal}}=-\frac{1}{N}\sum_{0\leqslant i\leqslant N}[y_i(1-p_i)^{\gamma}\ln p_i+(1-y_i)\,p_i^{\gamma}\ln(1-p_i)] \tag{8}$$

式中：参数γ是注意力参数，固定取值为 2；难分裂缝像素点的预测概率值p_i越小，所贡献的损失值就越大，模型就会更关注难分像素点的优化。

BCE 损失函数根据裂缝和背景像素点的分布情况将注意力集中于样本较少的裂缝像素，Focal 损失函数则根据所有像素点的预测准确度情况将注意力集中于预测不准的裂缝像素。可进一步为 Focal 函数添加权重因子：

$$L_{\mathrm{BFocal}}=-\frac{1}{N}\sum_{0\leqslant i\leqslant N}[w_1y_i(1-p_i)^{\gamma}\ln p_i+w_2(1-y_i)p_i^{\gamma}\ln(1-p_i)] \tag{9}$$

为研究哪种损失函数和权重比例在裂缝分割任务中更具优越性，本文进行了损失函数的预实验。固定w_2为 0.1，取不同的w_1对 2.3.1 节中 2 种平衡损失函数进行 4 组不同权重的实验，在验证集上的验证结果如表 2 所示。

表 2 不同权重的损失函数对比

Table 2 Comparison of loss functions with different weights

损失函数($w_2=0.1$)	aAcc	mAcc	*F*1	*P*	*R*
BCE($w_1=0.2$)	0.983 8	0.906 0	0.770 4	0.724 3	0.822 8
BCE($w_1=0.3$)	0.981 1	0.925 7	0.751 9	0.664 2	0.866 3
BCE($w_1=0.4$)	0.979 0	0.935 0	0.744 0	0.640 2	0.888 1
BCE($w_1=0.5$)	0.977 3	0.940 7	0.724 4	0.605 4	0.901 6
BFocal($w_1=0.2$)	0.984 8	0.865 8	0.763 3	0.790 0	0.738 4
BFocal($w_1=0.3$)	0.984 7	0.861 4	0.759 7	0.792 7	0.729 3
BFocal($w_1=0.4$)	0.984 8	0.864 7	0.761 8	0.789 3	0.736 1
BFocal($w_1=0.5$)	0.984 7	0.868 7	0.762 7	0.781 7	0.744 6

如表 2 所示，比例设置越高，召回率越高，而准确度越低。这是由于随注意力向裂缝元素偏移，模型更多地关注裂缝元素的分类准确度，所以裂缝元素被误判为非裂缝元素的可能性越小，召回率越高。与此同时，模型对非背景元素的关注度降低，导致准确率降低。当比例设置过高时，容易导致非背景元素误检为裂缝元素。

根据 *F*1 指标，2 种损失函数的权重参数均取$w_1=0.2$、$w_2=0.1$时得分最高，与另外 2 种不带权重的损失函数在验证集上的验证结果如图 5 所示。

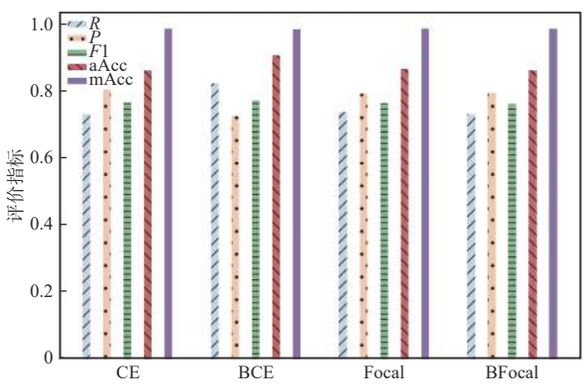

图 5 4 种损失函数对比

Fig. 5 Comparison of four loss functions

如图 5 所示, 采用 BCE 损失函数虽然降低了查准率 P, 但是查全率 R 提高, 使得模型更多地关注裂缝元素的预测, $F1$ 分数也随之略有提升, 整体上也具有更高的 aAcc 分数。因此本文采用裂缝像素点权重w_1为 0.2, 背景像素点权重w_2为 0.1 的平衡交叉熵损失函数。

3　实验结果和讨论

基于 2.1 节中的数据集, 本章对本文模型进行性能评估。包括探究改进策略对模型性能的影响, 评估在不同大小和覆盖度的数据集上模型的性能特别是泛化性能的表现, 并与其他分割模型 PSPNet、U-Net、FCN、DeepLabv3+进行定量和定性比较, 验证本文所提裂缝分割算法的有效性和实用性, 这些模型均采用 2.3 节中描述的训练方法和技巧。

3.1　消融实验

为了探究第 1 章所提改进策略的有效性, 从本文提出的 3 个改进策略为出发点进行了对比消融实验, 并将实验结果展示在表 3 中。实验在 UCrack 数据集上进行, 对比基线为编码器部分采用残差模块提取特征, 解码器部分保留上线性插值上采样, √代表采用, 为空代表未采用。

表 3　消融实验

Table 3　Ablation study

实验组	ResNet	ConvNeXt	DUC	PPM	aAcc/%	mAcc/%	$F1$/%	P/%	R/%
1	√				97.43	87.39	76.38	77.25	75.54
2		√			98.45	88.17	78.56	79.52	77.62
3		√	√		98.49	90.18	78.06	75.08	81.28
4		√		√	98.59	87.76	78.71	81.46	76.15
5		√	√	√	98.51	90.89	80.92	79.20	82.72

从表 3 可以得出, 3 种策略均可以提升裂缝分割的性能。仅使用 ConvNeXt 模块作为编码器的特征提取模块就能提高裂缝分割任务的整体性能表现, 这说明 ConvNeXt 模块在裂缝图像分割领域具有较强的优势。实验 3 在实验 2 的基础上验证了 DUC 模块的有效性, 检测召回率提高了 3.66%, 说明获取更多的细节信息能降低裂缝这种小目标的漏检率。实验 4 在实验 2 的基础上验证了 PPM 模块的有效性, 裂缝的检测准确度提高了 1.94%, 说明充分利用全局信息能降低噪声误检以提高精确率指标。实验 3 和实验 4 中性能提升的同时, 伴随着相对立指标的降低, 实验 5 验证了二者结合能够同时兼顾全局语义信息和局部细节信息, 从而使综合性能上 $F1$ 得分提高了 2.36%。

3.2　模型在不同数据集下的性能表现

从原训练集中分别随机挑选出三分之一、二分之一的数据作为新的训练数据集训练模型, 并在 UCrack 测试数据集和 CrackDataset 测试集上进行模型性能测试, 定量地比较不同大小的训练数据集对模型性能的影响。

表 4 所示是在本文测试数据集上的测试结果, $F1$ 分数从 76.54% 提高到 80.92%, 平均准确度均值 mAcc 从 86.95% 提高到 90.89%, 随着训练集数据量的增加, 模型的性能随之提高, 说明使用足够大的训练数据是有必要的。CrackDataset 数据集是独立于本文数据集 UCrack 的小样本数据, 表 5 是在 CrackDataset 上的泛化性测试结果。

表 4　所构建的 UCrack 数据集上的测试结果

Table 4　The test based on the established UCrack dataset

训练集大小/张	aAcc/%	mAcc/%	$F1$/%	P/%	R/%
959	98.49	86.95	76.54	78.59	74.60
1438	98.56	87.89	77.81	79.21	76.47
2876	98.51	90.89	80.92	79.20	82.72

表 5　CrackDataset 测试数据集上的测试结果

Table 5　The test results on the CrackDataset dataset

训练集大小/张	aAcc/%	mAcc/%	$F1$/%	P/%	R/%
959	96.99	90.44	79.23	75.92	82.84
1438	98.45	91.18	78.13	73.49	83.39
2876	97.05	91.79	80.11	75.22	85.68

如表 5 所示, 使用不同大小训练数据训练的模型在 CrackDataset 上的测试结果相差不大, 三分之一的训练数据和原训练数据在 $F1$ 分数上仅相差

0.88%。这说明本文模型从较少的训练数据中也可以学习到有效的裂缝特征，泛化到其他数据集亦能获取到较好的检测性能。

除了模型本身的学习能力较强之外，另一部分原因可能是本文构建的 UCrack 数据集场景覆盖度广，提供了丰富的特征供模型学习。为了验证数据集场景覆盖度的必要性，本文进行了跨数据集泛化性能实验。为了尽量控制训练集数量对模型的影响，对比仅使用三分之一 (959 张) 原 UCrack 训练数据集和 CrackDataset 训练数据集 (621 张) 的模型测试结果。其中 CrackDataset 是场景覆盖度低的单一数据集，UCrack 是本文构建的场景覆盖度高的数据集。

如表 6 所示，使用场景覆盖度高的 UCrack 训练出来的模型，在两组数据集上都取得了不错的测试精度，在 CrackDataset 上的跨数据集 $F1$ 测试得分比在 UCrack 自身测试集上高出 2.69%，mAcc 得分高出 3.49%，考虑到 UCrack 测试集的复杂度，这种反差是合理的。但是使用场景覆盖度低的 CrackDataset 训练模型，即使在自身数据集上能取得很高的分数，但是一旦在 UCrack 上进行跨数据集测试，仅获得 24.53% 的 $F1$ 分数，泛化性能差，无法应用到实际场景。

数据集的场景覆盖度和大小对模型的检测精确度和泛化性能有显著的影响，这说明为了解决复杂背景下分割难点问题，构建场景覆盖度高、数量足够大的数据集，并基于裂缝特征丰富多样的数据集进行路面裂缝分割模型的研究是非常必要的，这也是本文构建的 UCrack 数据集的原因。

表 6 场景覆盖度不同的训练数据集的测试结果

Table 6 The test results on training datasets with different scene coverage

训练数据集大小/张	测试集	aAcc//%	mAcc/%	$F1$/%	P/%	R/%
UCrack(959)	UCrack	98.49	86.95	76.54	78.59	74.60
UCrack(959)	CrackDataset	96.99	90.44	79.23	75.92	82.84
CrackDataset(621)	UCrack	86.41	74.95	24.53	15.26	62.63
CrackDataset(621)	CrackDataset	98.21	94.77	86.10	81.82	90.86

3.3 与其他模型的性能比较

为了验证本文模型的路面裂缝分割性能，与 4 种在裂缝分割领域有相关研究的语义分割模型 POSTNET、U-Net、FCN、DeepLabv3+进行比较，所有模型均基于相同的训练策略，使用相同的 UCrack 训练集进行训练。并在 UCrack 测试集上进行定量评估，同时对比所有模型在 CrackDataset 数据集上的泛化性能，评估结果如表 7 所示。

表 7 各模型在数据集上的测试结果

Table 7 The test results on the dataset of models

模型	测试数据集	aAcc//%	mAcc/%	$F1$/%	P/%	R/%	FPS
本文模型	UCrack	98.51	90.89	80.92	79.20	82.72	2.4
	CrackDataset	97.05	91.79	80.11	75.22	85.68	
FCN	UCrack	98.51	86.70	76.70	79.54	74.05	3.6
	CrackDataset	96.98	86.27	74.85	75.65	74.07	
U-Net	UCrack	98.52	87.33	77.10	78.92	75.35	3.4
	CrackDataset	97.42	89.01	78.87	78.29	79.45	
spent	UCrack	98.54	85.86	76.61	81.48	72.29	1.7
	CrackDataset	97.13	86.92	76.08	76.88	75.29	
DeepLabv3+	UCrack	98.60	87.61	78.24	80.80	75.83	3.1
	CrackDataset	97.35	87.87	77.93	78.8	77.08	

基于本文数据集的测试结果显示，本文模型虽然准确率指标分别略低于 PS Net 和 DeepLabv3+模型 2.28% 和 1.6%，但是召回率分别比 PSPNet 和 DeepLabv3+高 7.43% 和 6.89%，能够较好地实现精

确率降低不多和召回率提高较多两者之间的平衡, 即在不提高噪声误检的同时降低裂缝元素的漏检率。每秒传输帧数 (frames per second, FPS) 为每秒处理的图片数量, 用来评估模型的推理速度, 推理所用时间越短, 推理速度越快。虽然本文模型推理速度的表现属于中等水平, 但在 $F1$ 得分和平均准确度均值指标上取得较高值, 准确性优于其他模型。在 CrackDataset 上进行的跨数据集泛化性测试表明, 本文模型也取得了最佳效果, 相比表现最差的 FCN 在 $F1$ 分数上高 5.26%, 召回率高 11.61%, 相比其他在 $F1$ 分数上表现最好的 U-Net, $F1$ 得分高 1.24%, 但是召回率高 13.39%, 这验证了本文提出的裂缝分割方法具有更好的鲁棒性, 实际应用场景的适应性能力更强, 这得益于分层 ConvNeXt 特征提取网络提取到数据的多尺度特征, 经由特征金字塔结构多尺度进行特征融合和解码, 实现了较好的检测性能。

为了更直观地查看模型的检测效果, 基于尽量选取不同类型的测试样例的原则, 从裂缝尺度、光照条件、背景材料、污渍干扰等角度选取了不同测试样例进行可视化分析。所有可视化处理采用分割结果半透明化叠加原图进行输出, 不同行代表不同的测试样例, 不同列代表不同模型的测试结果, 其中 GT 列代表数据集自带分割信息的可视化结果。结果均使用半透明色填充矩形框标识检测结果, 其中实线框 (红色) 代表漏检情况, 虚线框 (黄色) 代表误检情况。如图 6 所示, 分别选取了细裂缝、粗裂缝、网状裂缝 3 种类型的 2 个样本。所有模型在粗裂缝样例检测中都只存在轻微的误差, 取得了不错的效果, 但是 FCN 和 PSPNet 模型在细裂缝样例上存在严重的漏检现象, DeepLabv3+和 U-Net 模型在网状裂缝的检测中也存在明显的漏检情况。虽然本文模型在裂缝细节上存在与其他模型一样的误检问题, 相对来说误检更多, 但是即使人眼也难以从细节上区分裂缝和噪声背景之间的准确界线。

此外, 选取了 7 种典型较复杂背景的难检测样例, 样例属于不同场景, 覆盖了水污、粗糙、椒盐噪点、车道线、花朵、修补等非裂缝干扰, 测试结果如图 7 所示。第 1 行和第 6 行的测试样例均包含污斑, 所有模型都能很好地识别; 第 2 行所示样例背景斑驳、裂缝细小, 人眼很难直接分辨出裂缝, 本文模型之外的其他模型在不同程度上出现了漏检问题; 第 3 行所示测试样例具有植物体和其产生的投影的干扰, 所有模型都能很好地区分植物体本身, 但其产生的投影边缘对 U-Net 检测产生干扰; 第 4 行所示样例裂缝中断部分的污斑与裂缝像

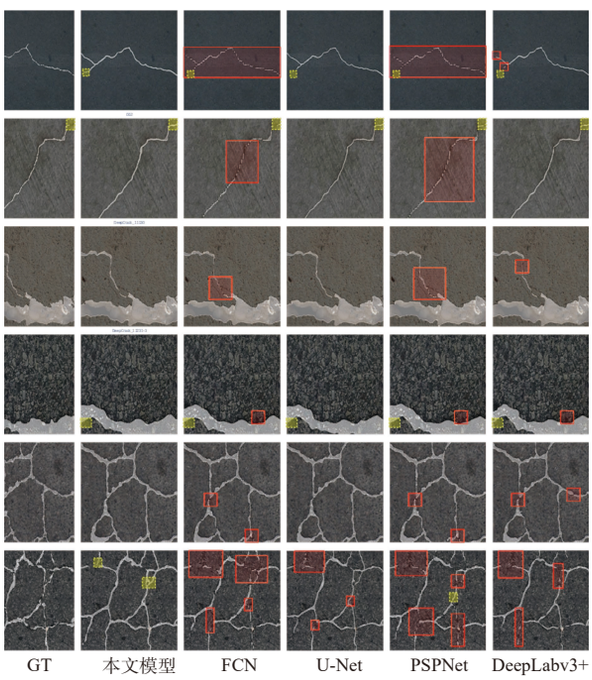

图 6　UCrack 中不同类型样例测试结果

Fig. 6　Test results of different types of data in UCrack

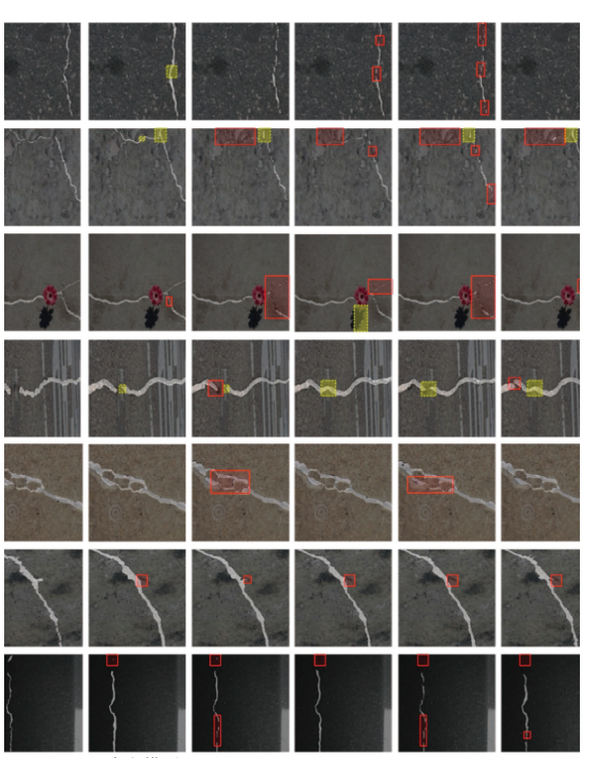

图 7　UCrack 中典型难分割样例测试结果

Fig. 7　Test results of typical hard-to-divide data in UCrack

素相似，所有模型都发生了误检现象，其中本文模型和 FCN 的误检率较低，但是 FCN 模型整体上存在较为严重的漏检问题；最后一行测试样例拍摄的光照较弱，所有模型都发生了漏检现象，其中本文模型和 U-Net 的漏检情况较轻。

3.4 泛化性定性测试

在 CRKWH100 上选取若干测试样例进行可视化分析，如图 8 所示。面对结构复杂的网状裂缝、光照度低、背景不均等难题，其他模型不同程度地表现出严重的漏检问题，本文模型总体上能较为准确地从背景中分割出裂缝像素。本文模型在 CRKWH100 上跨数据集泛化性测试的相对优秀的表现，说明本文方法可以较好地应对多场景的各种裂缝，具有良好的实际应用价值。

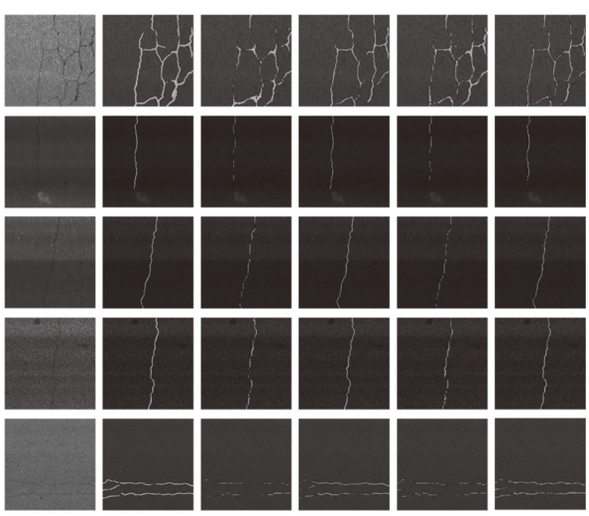

图 8 CRKWH100 上跨数据集检测结果

Fig. 8 Cross-dataset detection results on CRKWH100

除公开数据集，在 2.1 节所描述的南京路面裂缝图像上进行泛化性定性测试实验，若干测试结果如图 9 所示。第 1、2 列在椒盐噪声背景下整体的检测效果较好，只有边角的深色物形因形似裂缝而被误检；第 3、4 列表明模型能够较好地泛化于光照不均的场景，这 2 列也显示了模型在细裂缝上出现的轻度漏检现象，但裂缝的主要部分都能被很好地分割出来；第 5 列背景复杂，干扰和裂缝糅杂导致裂缝特征不明显，出现了较为严重的漏误检现象；第 6 列说明模型能够较好地分割出网状裂缝，一定程度上验证了模型对多尺度裂缝分割的泛化性；第 7 列在深色污渍上产生了误检，这是由于该污渍恰好在裂缝上，且颜色特征与裂缝十分相似而导致的。

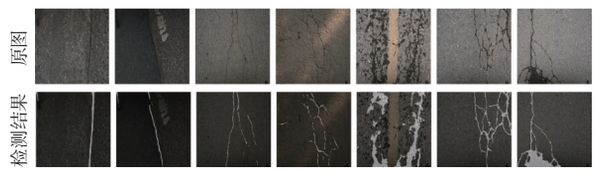

图 9 实际数据检测结果

Fig. 9 Detection results of actual data

4 结论

本文构建了一个在道路裂缝领域较大的数据集 UCrack，覆盖多种噪声干扰、裂缝类型和背景元素，用于复杂背景下提高裂缝分割准确度和模型泛化性能的研究。提出一种基于 ConvNeXt 卷积神经网络的分层结构的裂缝分割模型，与 FCN、U-Net、PSPNet 和 DeepLabv3+ 这 4 种先进的语义分割模型进行了对比评估，同时在公开数据集 CRKWH100 和作者在南京采集的数据集上进行了泛化性测试。结果证明，本文构建的数据集 UCrack 场景覆盖度高，能够提供丰富的裂缝特征；本文模型提取特征能力强，能够学习到足够的裂缝特征并拓展到更多的场景，具有优秀的泛化性能，能够有效应对复杂的实际应用问题。但是本文模型的误检问题相比其他模型没有得到改善，如何降低误检率是下一步研究的重点。

参考文献：

[1] 董元帅, 周绪利, 侯芸, 等. 基于寿命周期的沥青路面预养护时机决策优化[J]. 公路, 2020, 65(4): 325-331.
DONG Yuanshuai, ZHOU Xuli, HOU Yun, et al. Optimization of the asphalt pavement pre-maintenance timing decision based on life cycle[J]. Highway, 2020, 65(4): 325-331.

[2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[3] Anon. ImageNet[EB/OL]. [2022-11-19]. https://www.image-net.org/challenges/LSVRC/.

[4] YANG X, LI H, YU Y, et al. Automatic pixel-level crack detection and measurement using fully convolutional network[J]. Computer-Aided Civil and Infrastructure Engineering, 2018, 33(12): 1090-1109.

[5] CHOI W, CHA Y J. SDDNet: real-time crack segmentation[J]. IEEE Transactions on Industrial Electronics, 2020, 67(9): 8016-8025.

[6] 于海洋, 景鹏, 张文涛, 等. 基于残差和注意力机制的道路裂缝检测U-Net改进模型[J]. 计算机工程, 2023, 49(6): 265-273.
YU Haiyang JING Peng, ZHANG Wentao, et al. Improved U-Net model for road crack detection by combining residuals and attention[J]. Computer Engineering, 2023, 49(6): 265-273.

[7] JI A, XUE X, WANG Y, et al. An integrated approach to automatic pixel-level crack detection and quantification of asphalt pavement[J]. Automation in Construction, 2020, 114: 103176.

[8] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C/OL]//Proceedings of the European Conference on Computer Vision (ECCV). New York: IEEE, 2018: 801-818[2022-11-19]. https://openaccess.thecvf.com/content_ECCV_2018/html/Liang-Chieh_Chen_Encoder-Decoder_with_Atrous_ECCV_2018_paper.html.

[9] 廖延娜, 豆丹阳. 基于Mask RCNN的桥梁裂缝检测方法设计及研究[J]. 应用光学, 2022, 43(1): 100-105.
LIAO Yanna, DOU Danyang. Design and research of bridge cracks detection method based on Mask RCNN[J]. Journal of Applied Optics, 2022, 43(1): 100-105.

[10] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. New York: IEEE, 2017: 2117-2125.

[11] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 11976-11986.

[12] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2021: 9992-10002.

[13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. New York: IEEE, 2016: 770-778.

[14] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[M/OL]. New York: arXiv, 2017[2023-08-06]. http://arxiv.org/abs/1704.04861.

[15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[M/OL]. New York: arXiv, 2021[2022-11-19]. http://arxiv.org/abs/2010.11929.

[16] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2023-04-20]. http://www.arxiv.org/abs/1704.04861.

[17] BA J L, KIROS J R, HINTON G E. Layer normalization[EB/OL]. [2023-04-20]. https://arxiv.org/pdf/1607.06450v1.pdf.

[18] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). New York: IEEE, 2018: 1451-1460.

[19] XIAO T, LIU Y, ZHOU B, et al. Unified perceptual parsing for scene understanding[M/OL]. New York: arXiv, 2018[2022-11-16]. http://arxiv.org/abs/1807.10221.

[20] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 2881-2890.

[21] AMHAZ R, CHAMBON S, IDIER J, et al. Automatic crack detection on 2D pavement images: an algorithm based on minimal path selection[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 17(10): 2718-2729.

[22] LIU Y, YAO J, LU X, et al. DeepCrack: a deep hierarchical feature learning architecture for crack segmentation[J]. Neurocomputing, 2019, 338(21): 139-153.

[23] EISENBACH M, STRICKER R, SEICHTER D, et al. How to get pavement distress detection ready for deep learning? A systematic approach[C]//2017 international joint conference on neural networks (IJCNN). New York: IEEE, 2017: 2039-2047.

[24] LEI Z, YANG F, ZHANG D, et al. Road crack detection using deep convolutional neural network[C/OL]//IEEE International Conference on Image Processing. New York: IEEE, 2016[2022-11-19]. http://ieeexplore.ieee.org/document/7533052.

[25] YONG S, CUI L, QI Z, et al. Automatic road crack detection using random structured forests[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(12): 3434-3445.

[26] ZOU Q, ZHANG Z, LI Q, et al. DeepCrack: learning hierarchical convolutional features for crack detection[J].

IEEE Transactions on Image Processing, 2019, 28(3): 1498-1512.

[27] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[EB/OL]. (2017-11-14). [2023-04-20]. http://arxiv.org/abs/1711.05101v3.

[28] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.

[29] XIE S, TU Z. Holistically-nested edge detection[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015. 1395-1403.

[30] REN X X, XING Z C, XIA X, et al. Neural network-based detection of self-admitted technical debt: from performance to explainability[J]. ACM Transactions on Software Engineering and Methodology, 2019, 28(3): 15.1-15.45.

[31] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[EB/OL]. [2022-11-15]. https://arxiv.org/abs/1708.02002v2.